SEQUENTIAL MINIMAX SEARCH FOR A ZERO OF A CONVEX FUNCTION

O. Gross
S. M. Johnson

P-935 px

September 11, 1956

The RAND Corporation
1700 MAIN ST. • SANTA MONICA • CALIFORNIA

SUMMARY

Given: ~~We know initially~~ a positive and a negative value of a function at two given points respectively. The function is continuous and convex and is otherwise unknown but computable. Starting with this information, ~~we describe~~ a procedure is described for locating its unique root (on the starting interval) within an interval of minimum guaranteeable length in n steps, where a step consists of calculating the value of the function at any point we choose. The pertinent functional equation is derived and curves of the objective function are plotted for n = 1, 2, 3, 4 from data obtained from the Johnniac, RAND's Princeton-type high-speed digital computer. ( )

CONTENTS

1. INTRODUCTION

In the past, methods have been proposed and used for the computation of real roots of equations in a single variable. Notable among these is Newton's Tangent Method. However, this procedure is quite costly in many instances wherein the function whose zero is desired is inordinately complicated, and indeed, it is usually the case, polynomials and other simple functions excepted, that derivatives are more complicated to evaluate than the functions themselves. This last circumstance constitutes the principal motivation of later methods which are based solely upon evaluations of the functions themselves. We shall not describe any of these in detail here, however, but shall content ourselves with a single reference [1], and the remark that our method falls in this latter category. To our knowledge, none of these later methods are sequential minimax in character except for the bisection procedure and this in a restricted sense.

Let us now mention that due to personal preference, or otherwise established physical laws, the functions encountered in the past and presumably to be encountered in the future, had a strong tendency to be analytic in character. This happy circumstance coupled with the rare occurrence of the zero sought being a point of inflection of the graph of the function justifies the assertion that once a root is located on a sufficiently small interval by a positive and negative reading, the function will be inclined to be convex or concave throughout the interval.

Thus, for example, in [1] the equation $f(t) = t^3-6t^2+9t+6-6t^2e^{-t}-18te^{-t}-3te^{-2t}-6t^2 = 0$ is studied. $f(2)$ and $f(3)$ were found to be of opposite sign and thus a root was located on $(2,3)$. By repeated application of the "bisection" technique the root was found to be approximately 2.7184. However, to obtain 4 decimal accuracy with <u>certainty</u> using the bisection technique would almost always require (starting on an interval of unit length as above) 14 evaluations of the function even with the information that the function is monotone. On the other hand, it can be shown by elementary arguments that the above function f is convex on $[2,3]$. Making use of this information, together with the knowledge of $f(2)$ and $f(3)$ and succeeding functional values (the bisection procedure makes use only of the knowledge $(x_1, \operatorname{sgn} f(x_1))$ one should expect to obtain far more rapid convergence using a sequential minimax procedure which takes this information into account. This phenomenon will be exemplified in paragraph 3.

2. DESCRIPTION OF THE PROCEDURE

In what follows we shall describe a numerical procedure for solving the following problem:

"We know initially a positive and a negative value of a function at two given points respectively. The function is continuous and convex and is otherwise unknown but computable. Given an integer $n > 0$, how do we proceed to locate its root on this initial interval within an interval of minimum length in n steps where a step consists in calculating the value of the function at any point we choose?" As it stands, the question is meaningless, i.e., has no definite answer, until we specify, for example, that our procedure be sequential minimax, i.e., to say at each step of the procedure we assume that the worse possible situation might occur from that point on in the light of our present information about the function and proceed to evaluate the function at a point which hedges against all contingencies so as to guarantee a fixed interval at that stage.

An alternate question would envolve the final interval as specified in length and the requirement that the number of steps be minimized. Without going into any detail, we mention that the procedure which answers the first question need be somewhat modified to answer the second.

Finally we remark that if the function is concave rather than convex, we merely have to replace the function by its negative, i.e., $f(x) \rightarrow -f(x)$ for the procedure to be applicable. Also, if the signs of the initial readings are reversed, i.e., $a < b$ and $f(a) < 0$, $f(b) > 0$, where (a,b) is our starting interval, we need only replace $f(x)$ by $f(-x)$ to render the procedure applicable.

We now describe the procedure cycle:

Suppose we are in the situation in which we know $f(a) = Y_a > 0$, $f(b) = -Y_b$, $Y_b > 0$, where $a < b$, and the root $> S$ and we have n more readings to make.

Then we have bracketed the root on the interval (S,W), where

$$W = a + (b-a)\frac{Y_a}{Y_b+Y_b} .$$

If $n = 0$, the computation ceases and the values S, W are recorded.

If $n > 0$, calculate the value of f at $x = S + (W-S)\cdot\rho_n\left(\frac{Y_b}{Y_a}\right)$*.

If $f(x) = Y_x > 0$, set $a' = x$, $b' = b$ and $S' = x+(x-a)\dfrac{Y_x}{Y_a-Y_x}$.

If, however, $f(x) = -Y_x < 0$, set $a' = a$, $b' = x$, and if $Y_x > Y_b$ set $S' = S$, otherwise set $S' = \max(S, x-(b-x)\dfrac{Y_x}{Y_b-Y_x})$.

Finally, set $n' = n-1$.

Then we are in the situation in which we know $f(a') = Y_{a'} > 0$, $f(b') = -Y_{b'}, Y_{b'} > 0$, where $a' < b'$ and the root $> S'$ and we have n' more readings to make. This completes the cycle. (As the problem is stated, $S = a$ initially.)

In the next section we shall illustrate how this procedure works on a particular example.

Remark 1. The foregoing procedure is an approximation to the actual minimax procedure. The theoretically correct procedure would only involve replacing the expression $\rho_n\left(\frac{Y_b}{Y_a}\right)$ in the formula for x above by $\rho_n\left(\frac{S-a}{b-a}, \frac{Y_b}{Y_a}\right)$, where $\rho_n(S,Y)$ is defined in paragraph 4. Since the objective function is relatively insensitive to S in our choice of $\rho_n(S,Y)$ in the vicinity of the minmax, we feel that the approximation

$$\rho_n(0,Y) \approx \rho_n(S,Y)$$

is justified, and define $\rho_n(Y) \equiv \rho_n(0,Y)$. This approximation renders the procedure more adaptable to machine computation.

---

*Graphs of ρ_n for $n = 1,2,3,4$ are included at the end of this paper.

Remark 2. Since the average digital computer has difficulty in reading graphs, in order to program the procedure, one would first find suitable algebraic approximations to the graphs of ρ_n after the fashion of C. Hastings, Jr. [3].

3. A NUMERICAL APPLICATION—COMPARING THE BISECTION TECHNIQUE

Suppose we are desirous of bracketing in on the zero of a certain complicated function f defined over the interval (0,1). We know that f is continuous and convex and that in fact, $f(0) = 1$, $f(1) = -1$. However, since the function requires one hour of machine time to evaluate a single point, we are ignorant of the fact that to all intents and purposes it is given by the relatively innocuous expression $f(x) = \max(-1, (x - \frac{1}{3})(\frac{x}{2} - 3))$. We can afford to make three evaluations of the function ($n = 3$). Upon referring to the graph of $R_3(0,Y)$* with $Y = \frac{Y_0}{Y_1} = 1$ we see that we can guarantee locating the root on an interval of length .01 times the original interval, i.e., on an interval of length .01. However, since the graphs represent the worst that can happen to us, we expect to do much better, and indeed this turns out to be the case.

We proceed to calculate:

Cycle No. 1.

$a = 0$, $b = 1$, $Y_a = 1$, $Y_b = 1$, $S = 0$, $n = 3$, whence by our formula, $W = .5$, and we have located the root on $(0, .5)$. Next,

*At the end of this paper.

$x = 0 + (.5-0)\,\rho_a(1) = .5(.148) = .074$, and we find that $f(.074) = .76839 > 0$; so $a' = .074$, $b' = 1$ and $s' = .074 + \frac{.074(.76839)}{1 - .76839} = .31950$. Finally, $n' = 2$.

Cycle No. 2. (dropping primes on the new variables)

$a = .074$, $b = 1$, $Y_a = .76839$, $Y_b = 1$, $S = .31950$, $n = 2$, whence by our formula, $W = .074 + \frac{.926(.76839)}{1.76839} = .47636$, and we have located the root on $(.31950, .47636)$. Next, $x = .31950 + .15686\,\rho_a(\frac{1}{.76839}) = .31950 + (.15686)(.198) = .35066$, and we find that $f(.35066) = -.04795 < 0$, so $a' = .074$, $b' = .35066$ and since $Y_x < Y_b$, $S' = \max\left(.31950,\ .35066 - \frac{(1-.35066)(.04795)}{1 - .04795}\right) = .31950$. Finally, $n' = 1$.

Cycle No. 3.

$a = .074$, $b = .35066$, $Y_a = .76839$, $Y_b = .04795$, $S = .31950$, $n = 1$, whence by our formula, $W = .074 + \frac{(.35066-.074)(.76839)}{.76839 + .04795} = .33441$ and we have located the root on $(.31950, .33441)$. Next, $x = .31950 + (.33441-.31950)\,\rho_1(\frac{.04795}{.76839}) = .31950 + (.01491)\,\rho_1(.624) = .32440$, and find that $f(.32440) = .02535 > 0$, so $a' = .32440$, $b' = .35066$, and $S' = .32440 + \frac{(.32440-.074)(.02535)}{.76839-.02535} = .33294$. Finally, $n' = 0$.

Cycle No. 4.

$a = .32440$, $b = .35066$, $Y_a = .02535$, $Y_b = .04795$, $S = .33294$, $n = 0$, whence by our formula, $W = .32440 + \frac{(.35066-.32440)(.02535)}{.02535 + .04795} = .33348$, and we have located the root on $(.33294, .33348)$. $n = 0$, so the computation ceases and the interval $(.33294, .33348)$ is recorded.

Discussion

Since the bisection procedure does not take cognizance of the convexity of the function, one would obtain initially without any evaluation of f that the root lies on (0,1); with one evaluation, (0,.5); with two evaluations, (.25,.5); and finally, with three evaluations, (.25,.375). The lengths of these bracketing intervals are, then, 1, .5, .25, .125 respectively. Comparing these with those obtained by the procedure above, namely .5, .15686, .01491, .00054, we see that our procedure does have a slight edge in this instance. In fact, in any instance of a convex function with the same starting values as our example, we can guarantee in three evaluations, a bracketing interval of length $\leq .01$, as was pointed out earlier. This value compares favorably with .125.

4. DERIVATION OF THE FUNCTIONAL EQUATION

Given a continuous convex function f and two of its values, one positive and the other negative, (say $f(x_1) > 0$ and $f(x_2) < 0$ with $x_1 < x_2$) how should one search for the zero of the function which lies on the known interval (x_1, x_2)? Utilizing the principle of optimality of the theory of Dynamic Programming [2], we formulate the problem in terms of minimizing the maximum length of interval on which we can deduce that the zero is located in n readings taken sequentially.

First, by reduction to scale, we can always consider the following diagram:

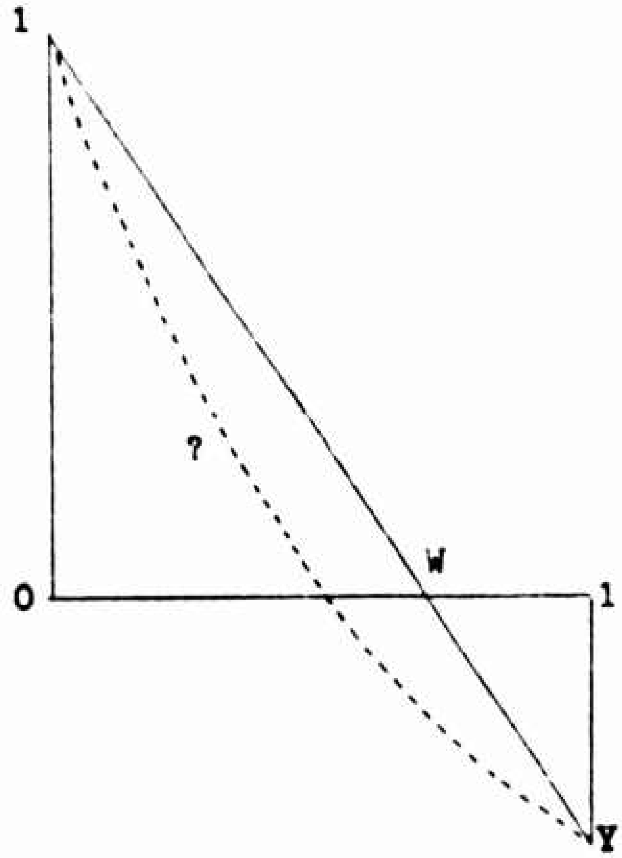


Fig. 1

where $f(0) = 1$, $f(1) = -Y$, $Y > 0$, $W = \frac{1}{1+Y}$. Since f is convex, the zero must lie on $(0,W]$. If we have just one more reading to make $(n=1)$, then we choose* x on $(0,W)$ and calculate $f(x)$. It can be shown by simple dominance arguments that no reading of f need ever be taken outside any interval on which the zero has been located. Having chosen x and calculated $f(x)$ we encounter exactly one of the following cases (barring $f(x) = 0$, of course, the best possible case).

<u>Case 1</u>

If $f(x) = v > 0$, then by drawing straight lines joining

---

*The particular optimal choice will be derived in what follows.

known points on the graph of f, we have the picture below, with the root located on $[S, W']$ as implied by the convexity of f.

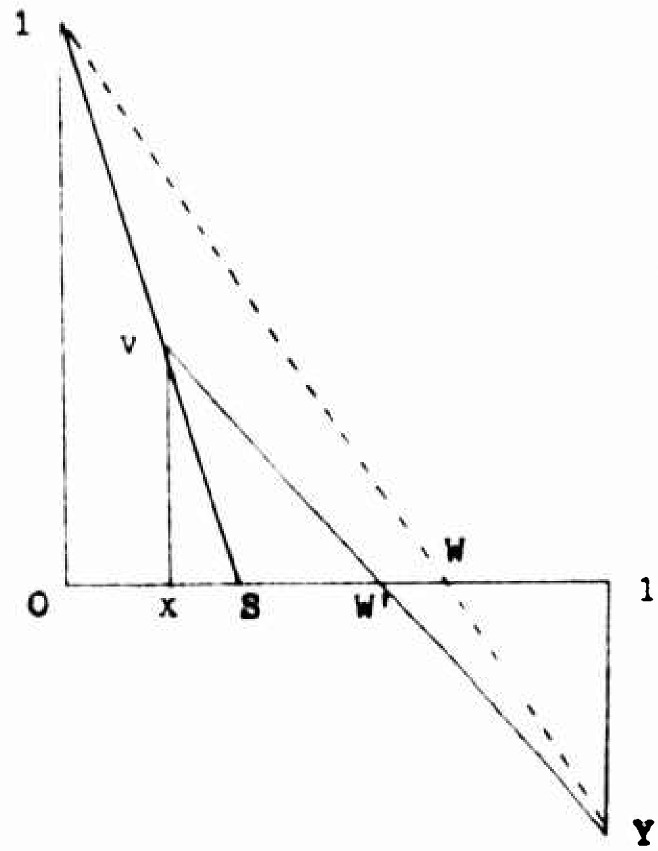


Fig. 2

Case 2

If $f(x) = -v' < 0$, the picture is

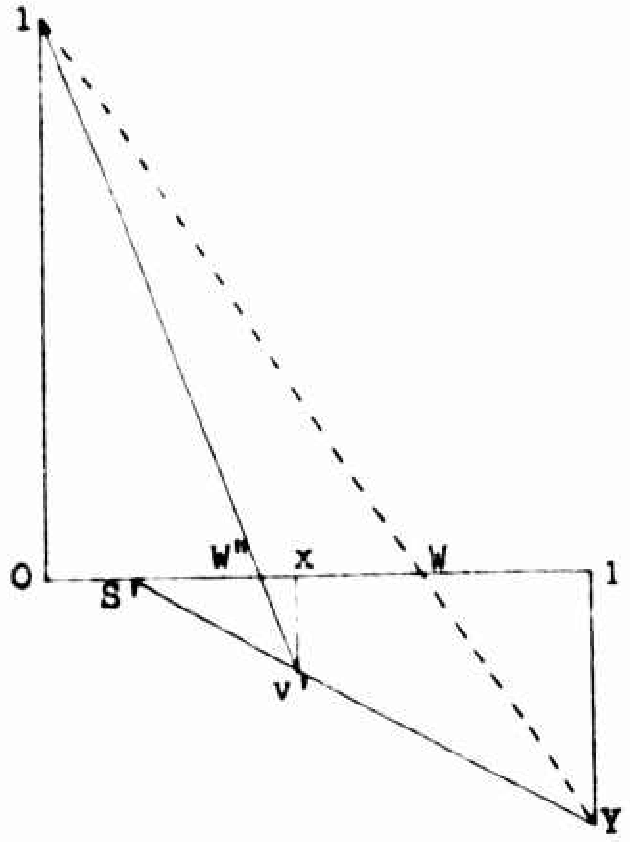


Fig. 3

with the root located on (max $(0,S')$, W'') as indicated.

Let $S'' = \max(0,S')$. If x is the final point at which $f(x)$ is to be determined (i.e. $n=1$) then it is chosen so as to minimize the maximum possible value of $\max(W'-S, W''-S'')$ consistent with our choice.

Before going into the algebraic details of the solution for $n=1$, let us consider the general n stage process in which we have several more readings to make. In either of the two cases diagrammed above, all the essential data can be described by a basic triangle determined by a two parameter system as follows.

In Fig. 2, we can conclude that the graph of f lies above the line segment vS and below the segment vW'Y (where the letters stand for both points and values in the obvious manner). If we now draw the line segment SY, it may or may not be crossed by the graph of f; but if it is crossed at some point P say, we can replace that portion of the curve lying below PY by PY with impunity. This device does not violate the convexity condition nor does it exclude the worst possible case. This can be shown by simple dominance arguments based on our knowledge of the function at any particular stage.

The essential data can thus be described by the triangle vSY. Since a vertical reduction in scale leaves the problem invariant, and a horizontal reduction, relatively invariant, the triangle vSY can be replaced by a triangle of standard form described pictorially by two parameters Y,S (say) as follows.

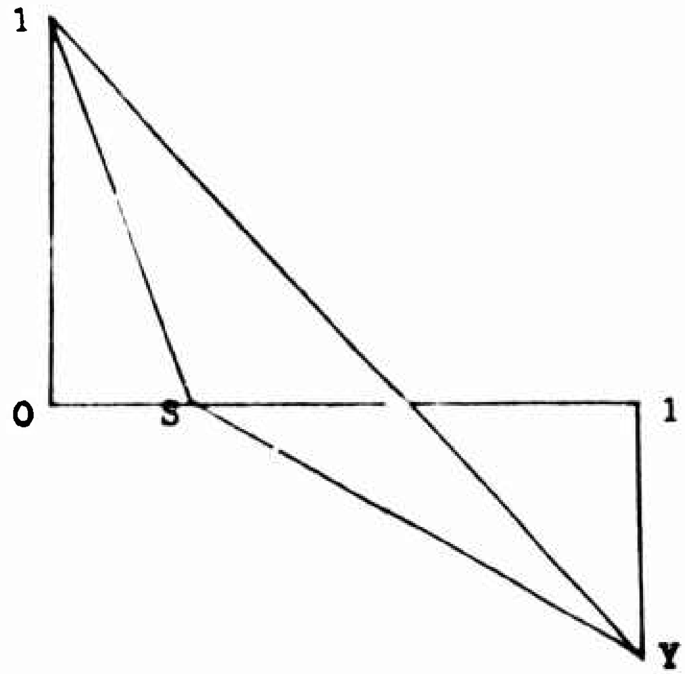


Fig. 4

Similarly, in the second case (Fig. 3), the graph of f lies above S'v' and below 1W"v'. Draw the line segment from 1 to S' and replace any portion of the graph of f lying below this line by the line itself from 1 to the point of crossing. This does not effect the choice of subsequent x's, since they will all be chosen on minimal bracketing intervals. Again, this can be shown by simple dominance arguments. Thus, in the second case also by a suitable reduction to scale we are led to another representation of Fig. 4.

Now define

$R_n(S,Y)$ = the minimum length of interval on which we can guarantee locating the zero in $[0,1]$ of any convex function f, given that $f(0) = 1$, $f(1) = -Y < 0$, the root is $> S$ and we have n readings to make. If $n = 0$, we have clearly,

$$F_0(S,Y) = \frac{1}{1+Y} - S \ .$$

Next, using the principle of optimality, and taking into account the scale factors, we obtain for $n > 0$ the following recurrence relation:

$$R_n(S,Y) = \min_{S \le x \le \frac{1}{1+Y}} \max \begin{cases} \displaystyle\max_{0 \le v' \le \frac{Y(x-s)}{1-S}} xR_{n-1}\left(\frac{xy-v'}{x(y-v')}, v'\right) \\ \\ \displaystyle\max_{0 \le v \le 1-x(1+Y)} (1-x)\, R_{n-1}\left(\frac{x}{1-x} \cdot \frac{v}{1-v}, \frac{y}{v}\right) \end{cases}$$

with the upper and lower expressions after the brace corresponding to the second and first cases, respectively. The scale factors are obtained in a completely elementary manner by means of similar triangles.

The ranges of the variables S,Y above are given by $Y \ge 0$, $0 \le S \le \frac{1}{1+Y}$. To render the expressions more amenable to computation on the Johnniac, we made the following substitutions:

$$W = \frac{1}{1+Y}\ , \quad \phi_n(S,W) = R_n(S,Y), \text{ whence } R_n(S,Y) = \phi_n\left(S,\frac{1}{1+Y}\right).$$

An additional modification of the quantified variables v, v' finally reduces the system to

$$\phi_0(S,W) = W-S$$

$$\phi_n(S,W) = \min_{S \le x \le W} \max \begin{cases} \displaystyle x \max_{S \le t \le x} \phi_{n-1}\left(\frac{t}{x}, \frac{W(1-t)}{W(1-x)+x-t}\right) \\ \\ \displaystyle (1-x) \max_{x \le t' \le W} \phi_{n-1}\left(\frac{t'-x}{1-x}, \frac{W(t'-x)}{t'-Wx}\right) \end{cases}$$

where $0 \le S \le W \le 1$.

The functions $\phi_n(S,W)$ were then computed for $n = 1,2,3,4$ by means of a discrete approximation using various grid sizes and linear interpolation. We are indebted to S. Dreyfus for programming these computations for the Johnniac.

The minimizing x's were of course recorded and these form the basis of our optimal policy, i.e. to say $x^* = x_n^*(S,W)$ is the point at which we evaluate our unknown function f given that the root lies on (S,W) in our basic triangle and there are n evaluations to be made. The point x_n^* is, of course, itself measured on the basic triangle. To take care of the general situation, we must of course relate our readings to the original scale. If we let $\rho_n(S,Y)$ denote the fraction of the distance between S and W occupied by x_n^*, we readily obtain

$$\rho_n(S,Y) = \frac{x_n^*(S,W)-S}{W-S} ,$$

where $W = \frac{1}{1+Y}$. It is now a relatively easy matter to relate x_n^* to the original scale and thus to obtain the procedure cycle outlined in paragraph 2.

Graphs of the functions $R_n(0,Y)$ and $\rho_n(0,Y)$ for $n = 1,2,3,4$ are included at the end of this paper.

5. REMARKS—THE SPECIAL CASE $n = 1$

We shall initiate this section with a few remarks intended primarily to validate certain assumptions made, tacitly or otherwise, in the derivation of the functional equation. We shall close this section with a brief treatment of the special case $n = 1$, $S = 0$, which provided us with an excellent check on the validity of the data obtained from the Johnniac. The remarks are as follows.

1. Suppose we are at a certain stage in an optimal sequential minimax search. We have several values of f computed and we are ready to choose our next point of evaluation. Let a,b denote the closest evaluation points on the left and right respectively of the minimal interval (S,W) on which the root is known to lie. Then in an optimal procedure, we choose x on (a,b). To see this, we need only state that since the unknown function f may indeed be piecewise linear (a possible contingency) outside (a,b) is is easy to see that any such subsequent reading would in such a contingency afford us no information regarding the character of f within (a,b) pertinent to the location of its zero, not already implied by the quantities a,b,f(a),f(b), S and W, or indeed by any of our previous readings for that matter. Similar but slightly more involved arguments can be given to show that the next reading should be taken on the interval (S,W).

2. In the treatment of Case 2, it was tacitly assumed that $Y_x < Y$, i.e., to say the SY line has a negative slope as shown in the figure. Again, it can be shown by dominance arguments that the worst situation occurs when f is monotone, and indeed when the graph of f lies above the line SY. It is this condition which determines the limits of the quantified variable v in the upper line of the functional equation for R_n. It is precisely such dominance considerations as these which, though enabling us to express the functional equation in a relatively simple form and to obtain an optimal "policy" therefrom via its recursive

computation, do not enable us to obtain an optimal "procedure" directly from the equation. However, all the happy contingencies, together with the information about the function f they afford are taken care of in the procedure outlined in paragraph 2. It is intuitively clear that the resulting procedure is sequential minimax.

3. We now conjecture that the functional equation for R_n may be validly simplified by assuming that the maximum in the upper line of the equation is always taken on at the upper end-point of the range of v. This is true for example if n=1. However, the authors were unable to prove the assertion in general and thus the transformed equation for ϕ_n was subsequently submitted to the Johnniac in its present form. Suffice it to say that the resulting data supported the conjecture.

4. $R_n(S,Y)$ is separately decreasing in S and Y. To see this for the first variable, for example, upon recalling the definition of $R_n(S,Y)$, we need only observe that, other things being the same, the information that the root is $> S$ <u>includes</u> the information that the root is $> S'$ if $S' < S$. On the basis of the additional information, then, we can clearly <u>guarantee at least</u> as short a final bracketing interval with a larger S value if we are proceeding optimally, i.e. $R_n(S,Y) \leq R_n(S',Y)$ if $S > S'$. An analogous argument applies for the second variable.

5. We conclude our remarks with a conjecture about the asymptotic behavior of $\rho_n(S,Y)$ as $Y \rightarrow \frac{1}{S} - 1$. In particular, we conjecture that there exists a positive number ρ_n^* such that

$\rho_n(0,Y) \downarrow \rho_n^*$ as $Y \to \infty$. This is easily shown to be true for $n = 1$, but the proof seems to be difficult for larger values of n. The precise value for $n = 1$ is known and is included in the accompanying graph together with "estimates" of ρ_n^* for $n = 2,3,4$.

We conclude this paper with a brief discussion of some results obtained for the case $n = 1$, $S = 0$. We shall spare the reader the elementary albeit complicated algebraic details involved and simply state that upon substituting the function

$$R_0(S,Y) = \frac{1}{1+Y} - S$$

in the right member of our functional equation and setting $S = 0$ in the result we obtain a relatively simple algebraic minimax problem for the determination of $R_1(0,Y)$.

The substitution $\rho = x(1+Y)$ then yields, upon performing the minimax operation, the following equation relating the optimal ratio ρ and the variable Y:

$$(\rho^4-2\rho^3-5\rho^2-2\ \ +1)Y^2 + 2(2\rho^3-\rho^2-3\rho+1)Y + (2\rho-1)^2 = 0 ,$$

with the restrictions $Y > 0$, $0 \leq \rho \leq 1$.

By some inexplicable quirk of fate the discriminant of the above quadratic in Y turns out to be precisely $8\rho^5$. Thus, we readily obtain the following rational parameterization of the (ρ,Y) curve:

$$\left.\begin{aligned} \rho &= \tfrac{1}{2}(1-t)^2 \\ Y &= \frac{4t^2}{1-4t-t^4} \end{aligned}\right\} \quad 0 \leq t \leq t_0 .$$

To determine the value of t_0 we note that the limiting form of the polynomial equation as $Y \to \infty$ is simply the coefficient of the leading term in Y set equal to zero:

$$\rho^4 - 2\rho^3 - 5\rho^2 - 2\rho + 1 = 0.$$

This equation has a unique root on the interval (0,1) and is given by

$$\rho_1^* = \frac{1+2\sqrt{2} - \sqrt{5+4\sqrt{2}}}{2} \approx .282 .$$

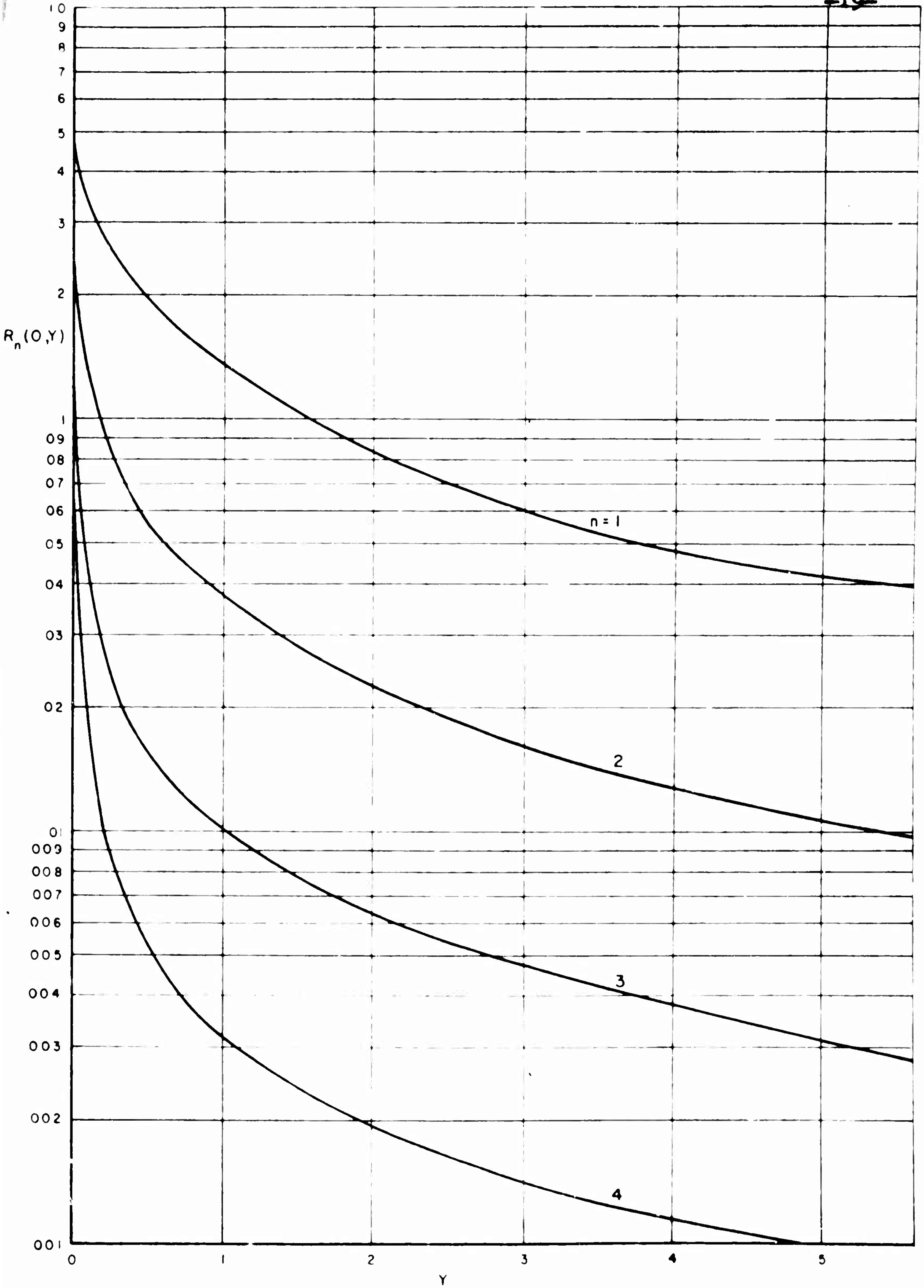
This is the asymptotic value mentioned earlier in remark 5. It follows from the parametric representation of ρ that $t_0 = 1 - \sqrt{2\rho_1^*} \approx .250$. As a check on this we note that this value of t_0 is the smallest positive root of the denominator in the parametric expression for Y.

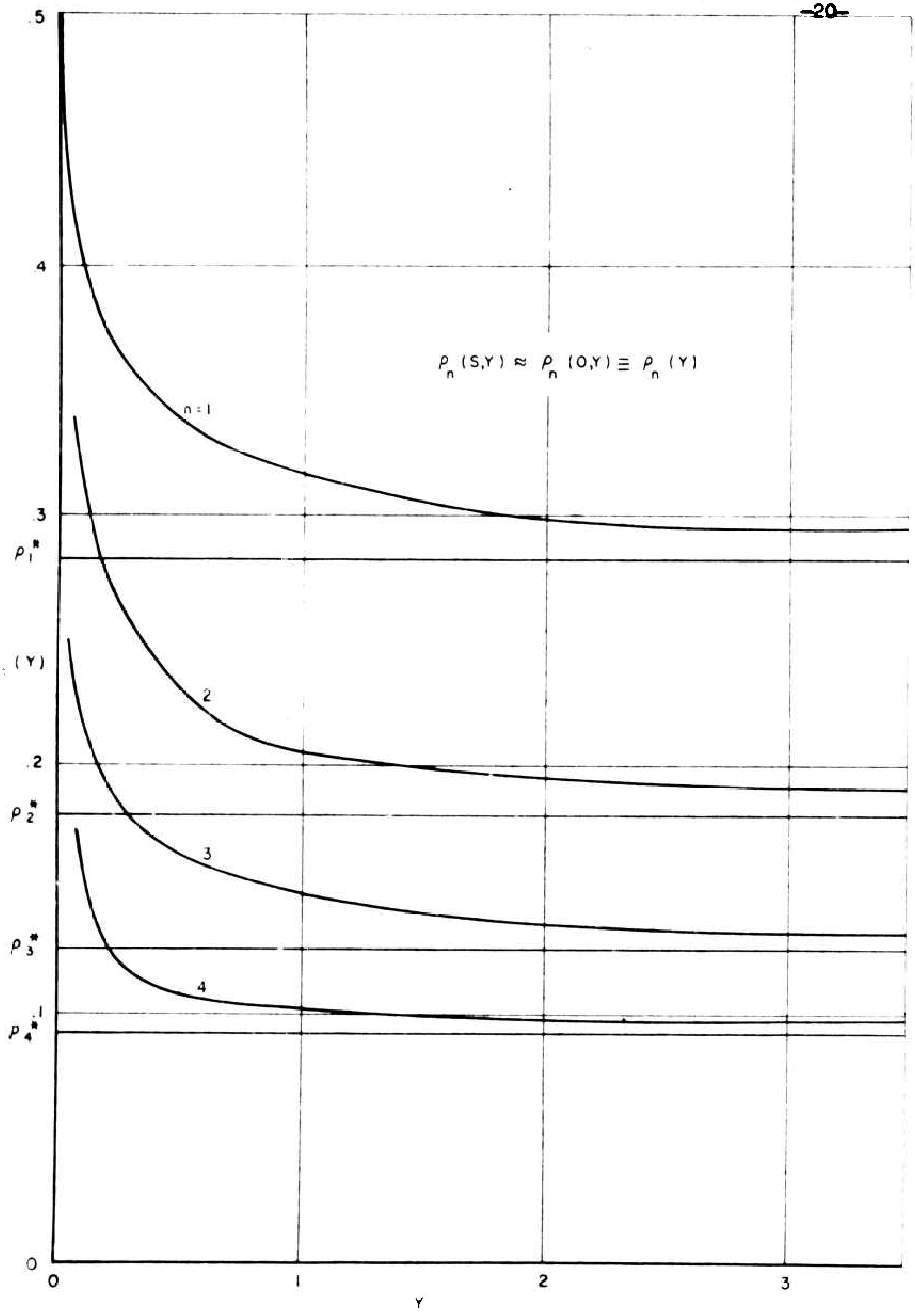
The graphs of $R_1(0,Y)$, $\rho_1(Y)$ were plotted from manual computations using the preceding formulas and these compared quite favorably with the results from the Johnniac. Unfortunately, however, due presumably to the choice of insufficiently small grid sizes imposed by the Johnniac's limited memory, a cumulative error caused, in effect, a gradual upward creep in the tails of the ρ curves. These were smoothed out to agree with theory to obtain the included graphs. The R curves, fortunately, seemed quite insensitive to choice of grid size.

In closing, we mention that further checks on the computations were provided by the easily derived relationships:

$$\phi_n(W,W) = 0, \quad \phi_n(S,1) = \frac{1-S}{2^n} ,$$

and these were found to fit the Johnniac data exactly.


10
9
8
7
6
5
4
3
2
$R_n(0,Y)$
1
.09
.08
.07
.06
.05
.04
.03
.02
.01
.009
.008
.007
.006
.005
.004
.003
.002
.001
n = 1
2
3
4
0
1
2
3
4
5
Y


.5
.4
.3
ρ_1^*
(Y)
.2
ρ_2^*
ρ_3^*
.1
ρ_4^*
0
0
1
2
3
Y
$\rho_n(S,Y) \approx \rho_n(0,Y) \equiv \rho_n(Y)$
n = 1
2
3
4

REFERENCES

[1] Shields, William H., Determination of Zeros of Polynomials and Real Roots of Transcendental Equations. Instrumentation Lab., Mass. Inst. of Tech. R-77, May 1954. 32p.

[2] Bellman, Richard, "The Theory of Dynamic Programming," Bull. American Math. Soc., Vol. 60, No. 6, pp. 503-515, November, 1954.

[3] Hastings, Cecil Jr. Approximations for Digital Computers, Princeton University Press, Princeton, N. J., 1955. 201p.